# The amplification of risk in experimental diffusion chains

Mehdi Moussaïd[a,b,1], Henry Brighton[b], and Wolfgang Gaissmaier[b,c,d]

[a]Center for Adaptive Rationality, [b]Center for Adaptive Behavior and Cognition, and [c]Harding Center for Risk Literacy, Max Planck Institute for Human Development, 14195 Berlin, Germany; and [d]Department of Psychology, Social Psychology and Decision Sciences, University of Konstanz, 78457 Konstanz, Germany

Understanding how people form and revise their perception of risk is central to designing efficient risk communication methods, eliciting risk awareness, and avoiding unnecessary anxiety among the public. However, public responses to hazardous events such as climate change, contagious outbreaks, and terrorist threats are complex and difficult-to-anticipate phenomena. Although many psychological factors influencing risk perception have been identified in the past, it remains unclear how perceptions of risk change when propagated from one person to another and what impact the repeated social transmission of perceived risk has at the population scale. Here, we study the social dynamics of risk perception by analyzing how messages detailing the benefits and harms of a controversial antibacterial agent undergo change when passed from one person to the next in 10-subject experimental diffusion chains. Our analyses show that when messages are propagated through the diffusion chains, they tend to become shorter, gradually inaccurate, and increasingly dissimilar between chains. In contrast, the perception of risk is propagated with higher fidelity due to participants manipulating messages to fit their preconceptions, thereby influencing the judgments of subsequent participants. Computer simulations implementing this simple influence mechanism show that small judgment biases tend to become more extreme, even when the injected message contradicts preconceived risk judgments. Our results provide quantitative insights into the social amplification of risk perception, and can help policy makers better anticipate and manage the public response to emerging threats.

risk perception | opinion dynamics | diffusion chains | social transmission | collective behavior

Public perception of risk is often polarized, difficult to anticipate, and at odds with scientific evidence (1, 2). The risks associated with nuclear energy, genetically modified food, and nanotechnologies continue to elicit strong public reaction in contrast to the assessments of many experts, and policy makers often fail to influence the public perception of the risks associated with poor nutrition, a sedentary lifestyle, and overexposure to the sun (2). The ability to communicate risks to the public, but also the ability to anticipate the public response to these risks, has a substantial impact on society and is key to alleviating unnecessary public anxiety (3).

Although research into the psychological factors that influence the formation of risk perceptions (4–7) and the communication of risks (8, 9) is relatively well developed, the study of risk perception is largely focused on the individual. However, risk judgments are formed in a social context: People frequently discuss the everyday risks they face with their friends, relatives, coworkers, and unknown people over the web. They observe and imitate the risky and self-protective behavior of their peers (10), and exchange opinions, sources of information, and behavioral recommendations through social media and online communication platforms. For example, during the H1N1 influenza outbreak in 2009, nearly half a million messages mentioning H1N1 vaccination were exchanged on Twitter, 20% of which explicitly expressed a positive or negative attitude toward vaccination (11). Indeed, a growing body of evidence suggests that people's perception of risk is mediated by social interaction. Large-scale social network analysis has shown that social interactions influence the spread of behaviors such as smoking,

food choice, and adherence to various health programs (12–14). In addition, it has been found that social proximity between individuals influences both their perception of risk (15, 16) and their emotional state (17).

These findings suggest that social interaction may lead to nonlinear amplification dynamics at the scale of the population. As illustrated by a number of studies of social contagion in relation to opinion dynamics (18), collective attention (19), cultural markets (20), and crowd behavior (21), local social influence can lead to chain reactions and amplification that yield population-scale collective patterns (22). Similarly, the dynamics of public risk perception bear the hallmarks of self-organized systems, such as opinion clustering (i.e., connected individuals share a similar judgment) and opinion polarization (i.e., opposing opinions coexist in the same population) (23). Therefore, a major challenge is to understand what role the social transmission of risk information between individuals plays in the formation of population-level dynamics. In particular, to what extent is risk perception contagious, and what behavioral patterns might this contagion yield at the macro level?

Few theorists have attempted to analyze and model these phenomena (24). Among those theorists who have, a common starting point is the social amplification of risk (SAR) framework, which attempts to combine the cultural, structural, and psychological factors that drive the formation of risk perception at the scale of the society (25, 26). The SAR framework suggests that individuals may play the role of "amplification stations" by transmitting a small and often biased subset of the available information. Although the overarching perspective of the SAR framework is widely assumed in risk research, it has undergone limited empirical analysis.

In this article, we study the social transmission of risk information from an interdisciplinary perspective by combining insights from the social and cognitive sciences with the study of

**Significance**

Public risk perception of hazardous events such as contagious outbreaks, terrorist attacks, and climate change are difficult-to-anticipate social phenomena. It is unclear how risk information will spread through social networks, how laypeople influence each other, and what social dynamics generate public opinion. We examine how messages detailing risks are transmitted from one person to another in experimental diffusion chains and how people influence each other as they propagate this information. Although the content of a message is gradually lost over repeated social transmissions, subjective perceptions of risk propagate and amplify due to social influence. These results provide quantitative insights into the public response to risk and the formation of often unnecessary fears and anxieties.

complex systems and nonlinear dynamics. In particular, we study how messages describing risk change in response to repeated social transmission by analyzing the relative fidelity with which two distinct aspects of a message—the message content and the message signal—propagate through human diffusion chains. What effect does the process of social transmission have on these message components, and what collective patterns of risk perception might this process generate?

To examine these questions, we analyze how information detailing the benefits and harms of the widely used but controversial antibacterial agent triclosan (27) is communicated from one individual to another in experimental diffusion chains. In a diffusion chain, a series of individuals propagate information sequentially and in turn from one individual to the next. Specifically, the information provided by the first individual is communicated to the second individual, who, in turn, communicates this information to a third individual, and so on. In our experiments, participants in the chain were instructed to communicate the risks surrounding triclosan in open, unstructured conversations. In each diffusion chain, the first participant was "seeded" with information presented in media articles detailing the benefits and harms of triclosan. In total, we examine 15 such diffusion chains, each composed of up to 10 participants. Both before and after participation in the experiment, we also assessed the subject's risk perception of triclosan. This experimental paradigm was pioneered by Frederic Bartlett over 80 y ago (28) and has since been used to study how a range of cultural entities undergo change, or cumulatively evolve, in response to repeated cultural transmission by humans and other animals (29–33). Here, we make use of the same experimental paradigm to study, for the first time to our knowledge, how risk information and, consequently, the risk perception of people change when socially transmitted. We found that subjects bias the signal of the message according to their subjective perception of risk, which influences the judgment of the receivers. Crucially, although risk perception "biases" propagate well and are typically amplified, the message content is transmitted with low fidelity and tends to become shorter, gradually inaccurate, and increasingly dissimilar between chains. We use computer simulations of this process to understand further the social implications of risk amplification. Put simply, although the content of messages describing risks degenerates in response to repeated social transmission, the signal of the message propagates with high fidelity, with social transmission playing the role of an amplifying process.

## Results

**Message Content.** We define the content of the message communicated from one participant to the next as the set of units of information that were communicated during a conversation between these participants. We identified 61 possible units of information by analyzing all conversations that occurred in the 15 diffusion chains. Each unit of information was classified in a three-level coding scheme (the detailed procedure is explained in *Materials and Methods*, with examples provided in Fig. 1) and tracked from one chain position to the next. This procedure allows us to study the propagation of information down each chain. In addition, a unit of information is labeled as "distorted" if it changes when transmitted from one chain position to the next. To illustrate, Fig. 1 depicts the propagation of risk information down a typical chain. At the first chain position, a total of 30 units of information were mentioned, 13 of which were propagated to the second chain position and only three of which were propagated to the final chain position. The three successfully transmitted units were, however, distorted (as illustrated by the color coding in Fig. 1). In addition, seven new units of information were generated as the chain unfolded (represented by squares in Fig. 1), two of which were propagated to the end of the chain. Propagation maps for all 15 chains are shown in Fig. S1. As shown in Fig. 2A, most units of information disappear as the chain unfolds, whereas those units that are propagated are done so with low fidelity and tend to become increasingly
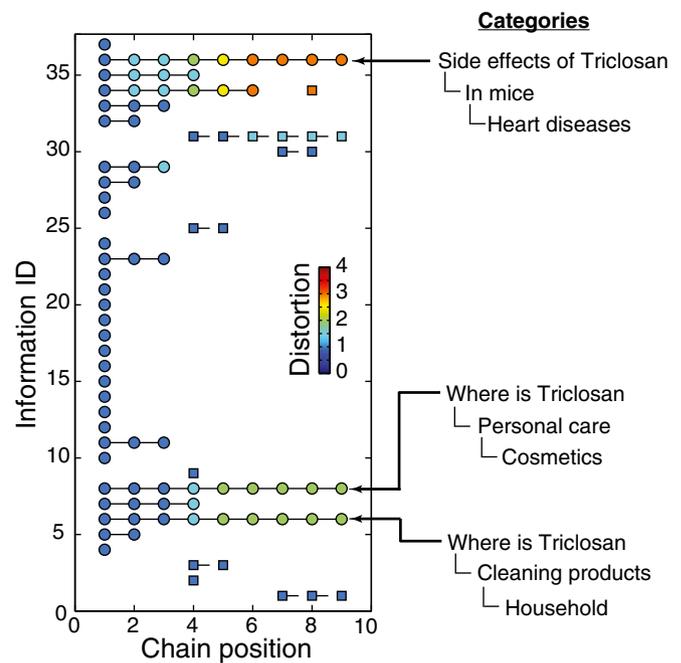


**Fig. 1.** Topological map of information propagation in an experimental diffusion chain. Among all units of information available at chain position 1 (blue dots), only three have survived to the end of the chain, although they were strongly distorted. The text on the right-hand side describes the categories of these units of information. Seven units of information were introduced by the chain (squares), two of which survived to the end of the chain. The color coding indicates the cumulated content distortion of the information. Information identifications (y axis) are arbitrary.

distorted. In addition, we measured the probability of a specific unit of information disappearing (pDeath) from one position to the next, as estimated from all 15 diffusion chains. As shown in Fig. 2B, pDeath has high values at the first two chain positions and reaches a relatively stable level of 0.2 afterward. Similarly, we estimated the probability that a specific unit of information is created (pBirth) or gets distorted (pDistortion) from one chain position to the next. As shown in Fig. 2B, these probabilities remain largely constant as a function of chain position. In addition, we found no significant differences when comparing these probabilities within each of the categories used in our coding scheme, which suggests that units of information (at least those units of information that we encounter here) can appear, disappear, or undergo content distortion with ostensibly constant probabilities, regardless of the kind of information communicated (Fig. S2). Next, we analyzed how the messages develop among the 15 chains. Do all of the chains eventually converge to a similar set of information units, or, conversely, do these sets diverge from each other? To examine these questions, we measured the message differentiation coefficient $D_{ij}^p$, which defines the proportion of information units present in chain $i$ at position $p$ that are not present in the message of chain $j$ at the same position $p$. Formally, the differentiation is defined as

$$D_{ij}^p = \frac{\left| m_i^p \notin m_j^p \right|}{\left| m_i^p \right|},$$

where $m_i^p$ is the set of information units contained in the message of chain $i$ at position $p$ and $\left| m_i^p \right|$ is the size of that set. Therefore, $D_{ij}^p = 0$ when the information units in chain $i$ are also present in chain $j$ at the same position (no differentiation), whereas $D_{ij}^p = 1$ indicates that none of the information units were found in chain $j$ at this position (maximum differentiation). Fig. 2C shows the
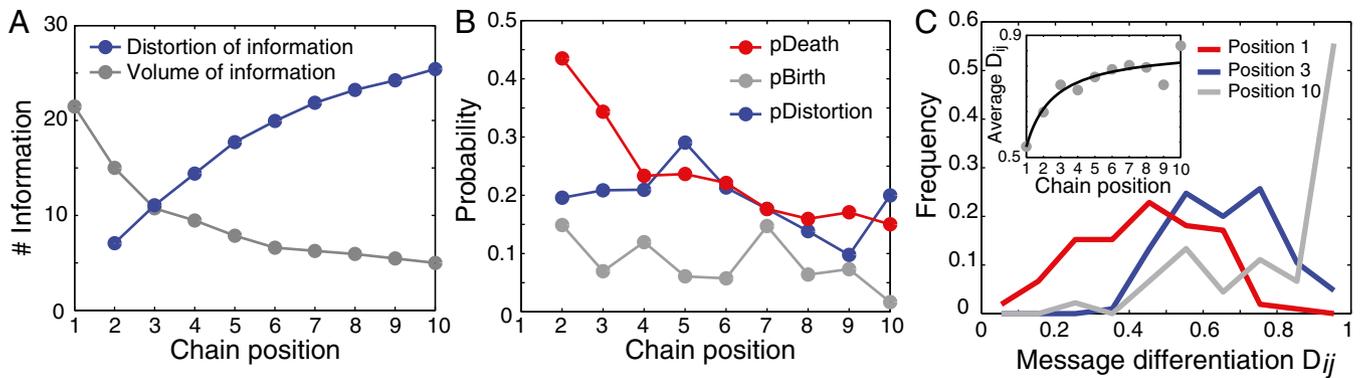
**Fig. 2.** Dynamics of information propagation. (*A*) Mean number of units of information transmitted and the mean cumulated content distortion of the information over the chains. (*B*) Hazard functions showing the probabilities that a piece of information disappears (red, pDeath), gets distorted (blue, pDistortion), or appears (gray, pBirth) at each chain position. (*C*) Distribution of the message differentiation $D_{ij}$ for all possible pairs of messages $i$ and $j$ at position 1 (red line), position 3 (blue line), and position 10 (gray line). The distributions tend to shift toward high differentiation values, indicating that the content of the message becomes increasingly different between the chains as it propagates from one person to another. (*Inset*) Mean value of $D_{ij}$ at each position of the chain (gray dots). The dark line corresponds to the fit equation $f(x) = p_1 x/(x + p_2)$ with parameters $p_1 = 0.86$ and $p_2 = 0.61$.

distribution of $D_{ij}^p$ for all possible pairs of messages $\{i,j\}$ taken at positions $p = 1$, $p = 3$, and $p = 10$. The distributions tend to shift toward high differentiation values as $p$ increases. Although the messages communicated at the first position have an average differentiation of 0.53, this value increases to 0.87 at the 10th position (Fig. 2*C*, *Inset*). In short, the content of the messages tends to become increasingly dissimilar among the chains as they propagate down each chain.

**Message Signal.** To understand further the impact of social information transmission, we analyzed the signal of the message, that is, whether the message carries a positive or negative assessment of triclosan. To quantify the signal of the message, we analyzed the transcripts of all conversations and marked each sentence as "negative," "positive," or "neutral" (*Materials and Methods*). Sentences are labeled negative when they express a negative assessment of triclosan (e.g., negative side effects) and are labeled positive when they express a positive assessment (e.g., that triclosan is safe). Sentences that express neither a positive nor negative assessment are neutral. This coding scheme is independent of the coding scheme used to quantify the message content. We use $n_p^+$ and $n_p^-$ to denote the number of positive and

negative statements found at chain position $p$ and $n_p$ as the total number of statements at position $p$. As shown in Fig. 3*A*, the three quantities decrease as a function of chain position but $n_p^+$ decays faster than $n_p^-$, suggesting that negative statements propagate down the chain more freely than positive statements. Consequently, the relative proportion of negative statements $\omega_p^- = n_p^-/(n_p^+ + n_p^-)$ tends to increase gradually, at the expense of the relative proportion of positive statements $\omega_p^+ = n_p^+/(n_p^+ + n_p^-)$ (Fig. 3*A*, *Inset*).

How are these aggregate trends related to the behavior of individual participants? We introduce the individual filtering coefficients $k_p^+$ and $k_p^-$ expressing the degree to which the participant at chain position $p$ modified the message (either positively or negatively) that he or she received from the participant at position $p-1$. The filtering coefficients are formally defined as $k_p^+ = n_p^+/n_{p-1}^+$ and $k_p^- = n_p^-/n_{p-1}^-$. Thus, $k_p^+ < 1$ holds for participants who attenuate the positive aspect of the message, whereas $k_p^+ > 1$ holds when a participant has amplified the positive aspect of the message (for negative statements, we have $k_p^-$, respectively). The distribution of $k_p^+$ and $k_p^-$, across all participants, is shown in Fig. 3*B*. In line with our previous finding, $k_p^-$ is, on average, higher than $k_p^+$ ($P < 0.001$), which further underscores that our participants
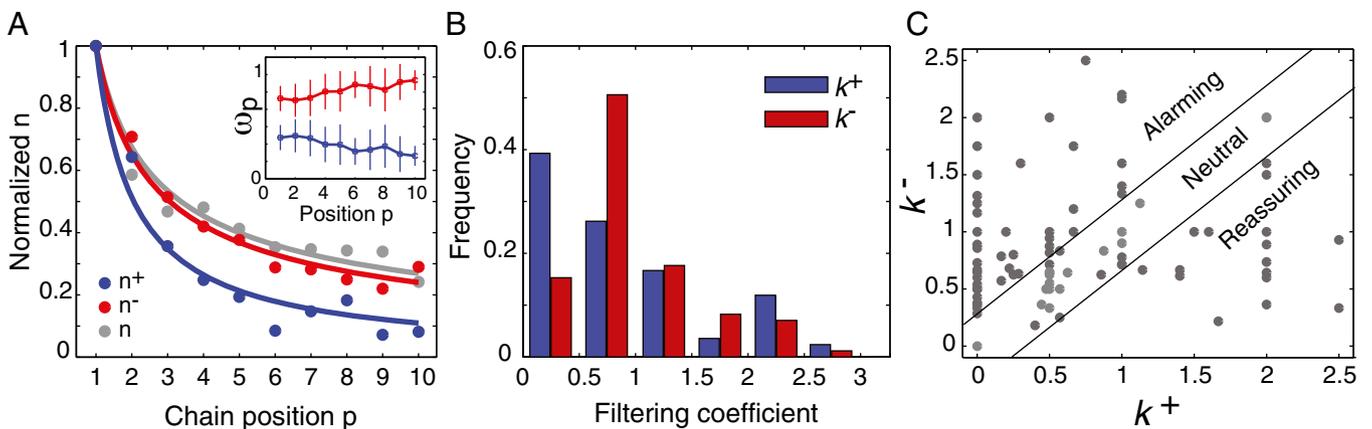


**Fig. 3.** Message mutation. (*A*) Evolution of the normalized number of positive statements $n_p^+$ (blue), negative statements $n_p^-$ (red), and total statements $n_p$ (gray) over the chain. Fit lines are power functions $f(x) = x^{-e}$, where the exponent $e$ equals 0.96, 0.62, and 0.57 for $n_p^+$, $n_p^-$, and $n_p$, respectively. (*Inset*) Average proportion of positive statements ($\omega_p^+$, in blue) and negative statements ($\omega_p^-$, in red) along the chains. (*B*) Distribution of the filtering coefficients $k_p^+$ and $k_p^-$ for all subjects. The distribution of $k_p^-$ is significantly higher than the distribution of $k_p^+$ ($P < 0.001$). (*C*) Individual profiles measured as the pair $\{k_p^+, k_p^-\}$. Each point represents one experimental subject. Individuals with $k_p^+ \approx k_p^-$ have a neutral effect on the message, whereas individuals with $k_p^- > k_p^+$ (respectively $k_p^+ > k_p^-$) tend to make the message more alarming (respectively more reassuring).

Moussaïd et al.

have an overall tendency to amplify negative statements and attenuate positive ones. However, significant individual differences were found (Fig. 3C): Although the majority of subjects tend to amplify the harmful aspects of triclosan ($k_p^- > k_p^+$), other participants have the opposite profile ($k_p^+ > k_p^-$) or act neutrally on the message ($k_p^+ \approx k_p^-$). Interestingly, the individual differences can be partly explained by the risk perception $\alpha$ that subjects reported in questionnaires conducted after the experiment [$\alpha$ correlates positively with $k_p^- - k_p^+$; correlation (c) = 0.25, $P = 0.019$]. In short, the direction of the "mutation" of the message corresponds to the opinion of the speaker: Individuals with higher risk perception tend to filter out positive statements and amplify the dangerous aspects of triclosan, whereas those individuals with lower risk perception tend to have an opposite effect on the signal of the message.

Finally, we measured what impact a message has on the risk perception of the receiver. Overall, participants changed their risk perception by an absolute average value of $|\Delta_\alpha| = |\alpha - \alpha^0| = 0.19$ (SD = 0.17), where $\alpha^0$ is the risk perception of the individual before the experiment and $\alpha$ is the risk perception of the same subject after the experiment. The degree of change $\Delta_\alpha$ should be interpreted relative to the signal $\sigma_p$ of the message that this participant received, defined as $\sigma_p = \omega_p^-$. When the message signal confirms the prior assessment of the participant (i.e., $\sigma_p \approx \alpha_0$), no change of opinion is expected (i.e., $\Delta_\alpha \approx 0$), whereas $\Delta_\alpha$ should vary in the same direction as $\sigma_p - \alpha_0$. As shown in Fig. 4, $\Delta_\alpha$ and $\sigma_p - \alpha_0$ correlate well (c = 0.55, $P < 0.001$), confirming that participants are influenced by the signal of the message they receive.

**Collective Patterns.** These findings provide quantitative insights into the role of social transmission as an amplifying process. Specifically, the signal of the incoming message influences the receiver's risk perception, which, in turn, shapes the signal of the outgoing message. In many social systems, similar feedback loops are typically associated with nonlinear dynamics and emerging collective patterns (22). In our experiment, the social amplification of the risk signal is clearly visible (Fig. 3A). However, considerable behavioral fluctuations are observed within and between chains, making it difficult to conduct a systematically controlled analysis of the amplification process. We analyzed a simple simulation model derived from our observations to study this amplification process more precisely.

In the model, each individual $i$ in a chain is has an initial risk perception $\alpha_i^0$. Individuals sequentially, and in turn, receive and transmit a message made of $n^+$ and $n^-$ positive and negative statements, respectively. The signal $\sigma$ of the message is defined as above, by the equation $\sigma = n^- / (n^+ + n^-)$. When an individual $i$ receives the message, the risk level of that individual changes as follows:

$$\alpha_i = \alpha_i^0 + s \left( \sigma - \alpha_i^0 \right), \qquad [1]$$

where $s$ is an influence factor with a value between 0 (no influence) and 1 (full adoption). Previous empirical estimates of influence factors lie between 0.3 and 0.5 (34, 35), which is consistent with our experimental estimate ($s = 0.45$; Fig. 4). In our simulation, when a message is transmitted from one person to another, it undergoes a random biased mutation dependent on the risk perception $\alpha_i$ of the emitter $i$. For the sake of simplicity, we implement the message mutation such that $n^+$ is increased by 1 with a probability $p_i^+ = \alpha_i$, and is decreased by 1 with a probability $p_i^- = 1 - \alpha_i$. Similarly, $n^-$ is increased by 1 with a probability $p_i^-$, and is decreased by 1 with a probability $p_i^+$. Based on this simple model, we implemented a series of computer simulations. Each simulation models a diffusion chain of $N$ agents, where the $i$th agent receives a risk message from the $(i - 1)$th agent and transmits a message to the $(i + 1)$th agent. Each agent $i$ has an initial risk level of $\alpha_i^0$. The first agent, at chain position $i = 1$, receives a seeded message composed of $n^+$ positive statements and $n^-$ negative statements. Iteratively, and for each agent in turn, three operations are applied: (i) agent $i$ updates his or her risk perception in response to the received message, as defined by Eq. 1, (ii) agent $i$ modifies the message using probabilities $p_i^+$ and $p_i^-$, and (iii) agent $i$ transmits the modified message to agent $i + 1$. Three example simulations illustrating the implications of different parameter values are detailed in Fig. S3.

Whereas the impact of the message on the first agent in the chain is given by Eq. 1, the dynamics at subsequent chain positions are complex and result from repeated mutations of the message, combined with related changes in the judgment of the individuals. First, we conducted simulations under conditions matching the conditions of our experiment, where there exists a substantial diversity of initial risk perceptions (i.e., where $\alpha_i^0$ is drawn from a normal distribution with a mean of 0.65 and SD of 0.22). At chain position $p = 10$, the aggregate distributions of risk signals and risk perceptions correspond to the risk signals and risk perceptions we observed experimentally (Kolmogorov–Smirnov test: $P = 0.57$ and $P = 0.08$). As we observed previously, the model predicts a gradual amplification of the risk signal, whereas individuals' final risk perception exhibits a considerable variability due to the diversity of initial judgments (simulation examples in Fig. S3). Under homogeneous initial conditions, where all individuals have the same initial risk perception (i.e., $\alpha_i^0 = \alpha^0$ for all individuals $i$), the amplification process is itself amplified and easier to study. For example, our simulations show that a neutral message (i.e., $n^+ = n^-$) injected in a population of concerned individuals (i.e., $\alpha^0 = 1$ for all individuals in the chain) tends to mutate rapidly as it propagates from one agent to another, culminating in a steady state representing the population's initial view, with no impact on the individuals' opinions in the long run (Fig. S3B). What social dynamics emerge when the initial risk perception $\alpha^0$ of the group, the signal $\sigma$ of the injected message, and the social influence factor $s$ are varied? Fig. 5 and Fig. S4 examine such an exploration of the model's parameter space. In short, the social amplification of the risk signal is visible after 10 transmissions and continues to increase as the message propagates further (Fig. 5A). After 50 transmissions, the signal of the message is either strongly positive (dark blue, Fig. 5A) or strongly negative (dark red, Fig. 5A). The impact on individuals' risk perception follows a similar trend (Fig. 5B). In most cases, the message has a polarizing effect on the population, where the biased initial judgments of the agents tend to become more extreme at the end of the chain, even when the injected message is
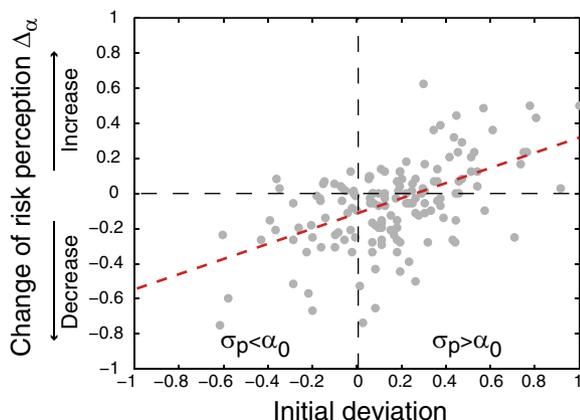


**Fig. 4.** Social influence. The scatter plot represents the degree to which participants changed their risk perception as a function of the initial deviation with the message signal they have received. Subjects receiving a message with negative signal, with respect to their initial perception, tend to increase their risk level, whereas those subjects receiving a more positive signal tend to reduce their risk level. The equation of the fit line is $y = 0.45x - 0.14$.

completely neutral. Along the transition zone depicted in green (Fig. 5), the mean values of the signal and the risk perception are both 0.5, but these means have very high SDs, indicating a U-shaped distribution where the outcome can be either high or low (Fig. S5). Thus, even a neutral message injected into a neutral population can possibly have a polarizing effect on the judgment of the individuals due to the amplification of random mutations of the message (as illustrated in Fig. S3C).

## Discussion

Among the factors influencing public risk perception, the impact of social transmission is arguably the least studied, despite being an everyday occurrence in an increasingly connected society. This study is the first, to our knowledge, to examine experimentally how the social transmission of risk information among human subjects influences the propagation of risky information, and how these risk perceptions become amplified as a result of this propagation. Ambiguity surrounds the risks associated with triclosan, such that each message not only carries factual information about the potential benefits and harms but also carries a signal representing the subjective judgment of the communicator. Consequently, we have distinguished between "content" propagation and "signal" propagation. First, we found that the information content of a message degrades (contains fewer units information) and becomes less accurate (undergoes content distortion) in response to being transmitted. As a consequence, different communication chains led to a focus on different issues related to triclosan, including Greenpeace protests (chain 13), breastfeeding (chain 10), and environmental damage (chain 3). Participants nevertheless influence each other: Changes in risk perception are a function of the signal of the received message (whether the message is positive or negative). Thus, changes in risk perception occur less as a result of the message content and more as a result of its overarching, subjective signal. In agreement with Bartlett's experiments (28), we observed that the information passed along the chain is shaped by the preconceptions of its members. In Bartlett's experiment (28), a Native American story was gradually modified to fit the cultural perceptions of the British participants. In our experiment, a similar mechanism drives the propagation of, and modifications to, "stories" detailing the risks associated with triclosan. One explanation for this behavior is that when faced with conflicting information, participants are in a state of cognitive dissonance and, consistent with Festinger's theory (36), reduce the discrepancy between their prior judgment and the received message. Thus, cognitive dissonance theory may partly explain why participants alter the strength of the risk signal.

At the aggregate level, the process of social transmission tends to amplify or attenuate the message signal. As a message propagates down the chain, it becomes distorted to fit the view of those individuals transmitting it, with the original message eventually having a negligible impact on the judgment of the participants. Ultimately, the message can have a counterintuitive polarizing effect on the population: After being transmitted several times, the message can strengthen the existing bias of the group, even though it initially supported the opposite view. As our simulations show, this social phenomenon is itself amplified in chains of like-minded individuals. According to recent theoretical and empirical studies (15, 23, 37), social interactions and information exchange frequently take place within clusters of like-minded people, suggesting that "natural" social structures may create favorable conditions for strong amplification dynamics. Investigating the development of risk patterns in naturally occurring social networks, which includes the impact of complex topologies with highly connected individuals (38) and a possibly stronger social influence (Fig. S4), therefore remains an important direction for future work. Similarly, factors influencing the frequency with which people communicate risks may also regulate their amplification. For example, new incidents typically lead to a burst of social exchange followed by a long-tailed decay of collective attention (19, 39).

More generally, our experiments demonstrate how the field of opinion dynamics can be extended to include the study of risk perception and how existing theoretical models of collective risk perception can be furnished with empirical support (37, 40). Perhaps most importantly, our findings illustrate the importance of studying patterns of risk perception as, at least in part, the outcome of a social process. From a public health perspective, the social amplification of risk can have undesirable and costly consequences, making it crucial to understand how policy makers can communicate risks in such a way as to aid their transmission through social networks. Promising lines of inquiry include simple yet effective methods for helping people assess the risks and benefits of technologies, chemicals, or medical drugs (8), and simple tabular representations and visual displays used to clarify public debates (41).

## Materials and Methods

**Experimental Design.** The experiment took place in April 2013. We invited 12 participants to each of the 15 experimental sessions. Due to absences, some sessions were conducted with less than 12 participants. In total, we collected data from four chains of length 12, two chains of length 11, four chains of length 10, three chains of length 9, one chain of length 8, and one chain of length 7 (average chain size = 10.1). Our analyses considered only the first 10 participants of diffusion chains of length greater than 10. All participants gave informed consent to the experimental procedure and received a flat fee of €15. The participants entered a waiting room and were instructed not to interact with each other. All participants completed an initial questionnaire Q1. The first participant $p_1$ was then moved to the experimental room and instructed to read a collection of six media articles displayed on a computer screen. The articles presented alternative views on the use and suspected side effects of the controversial antibacterial agent triclosan (these articles are provided in SI Appendix). To present a representative selection of articles, and one that reflects those articles likely to be encountered in an everyday setting, we selected each article from the first page of results returned by Google with the search keyword "triclosan" (accessed early in 2013). The articles were presented in a random order across groups, and subject $p_1$ had 3 min to read each of them. After the reading phase (18 min in total), the computer was shut down and participant $p_2$ was invited to join participant $p_1$ in the experimental room. Both participants were instructed to discuss triclosan in an open, unstructured discussion. No time limit was imposed. At the end of the discussion, the next participant $p_3$ was moved to the experimental room and a new discussion began between
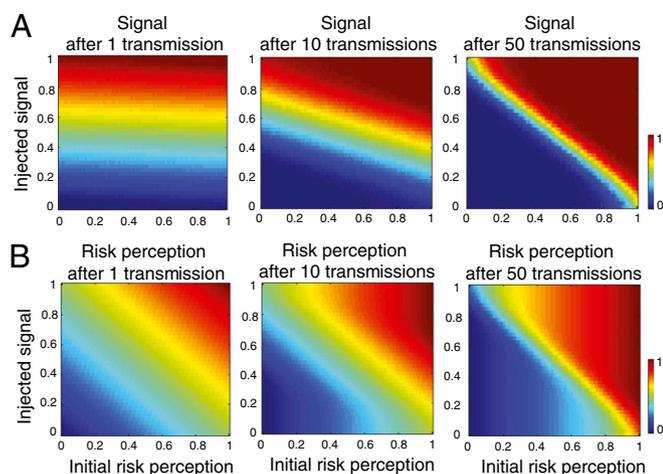
**Fig. 5.** Computer simulations. (A) Evolution of the risk signal at chain positions 1, 10, and 50 as a function of the initial risk perception of the individuals (x axes) and the signal of the injected message (y axes). The gradual dominance of extreme values (in dark red and dark blue) demonstrates the amplification of the risk signal. (B) Evolution of individuals' risk perception follows a similar trend. Individuals who initially expressed extreme risk perception (x = 0 or x = 1) gradually move back to their initial view, regardless of the message signal. The social influence parameter is set to s = 0.5. Simulations varying the value of s are provided in Fig. S4. Results are averaged over 500 simulations.

participants $p_2$ and $p_3$ under the same conditions. In the meantime, the first participant $p_1$ was instructed to complete the second questionnaire Q2 in another room and was then free to leave. This procedure was repeated for all participants in the chain. The discussions were recorded, and the audio files were then transcribed and translated into English (from German).

**Measures of Risk Perception.** The risk perception of the participants toward triclosan was measured before and after the discussion by means of two questionnaires Q1 and Q2. Questionnaire Q1 evaluated the subject's knowledge about triclosan. Among all participants, only one had prior knowledge of triclosan but failed to answer basic questions relating to triclosan. We therefore considered all subjects to be unfamiliar with triclosan. Questionnaires Q1 and Q2 also evaluated the risk perception of the participants toward triclosan. The risk perception of the participants was self-reported by placing a tick on a continuous line ranging from "Not dangerous at all" to "Extremely dangerous" (coded correspondingly with values between 0 and 1). Because participants were unfamiliar with triclosan before the experiment, it was impossible to assess their level of risk perception directly. We therefore used an indirect measurement by evaluating their perception of risk toward the more general issue of chemical use in food safety and cosmetics.

**Coding the Content of the Message.** We first used the transcripts of the discussions to identify every unit of information that was discussed. We identified a list of 61 units of information. These units were classified into four high-level categories: "Side effects," "Personal anecdotes," "Where is triclosan," and "Others." Each of these high-level categories was then subdivided into subcategories and sub-subcategories. For example, the top category, Side effects, contains the subcategory "In mice," which was further divided into seven classes, such as "Heart diseases" and "Cancer." Finally, we analyzed the transcripts and tracked each unit of information from one chain position to the next until the end of each chain. In addition, each unit of information

was labeled distorted if factual differences or imprecisions were detected from one chain position to the next. A content distortion was marked when (i) a numerical value has changed or disappeared; (ii) a qualitative indication of volume, frequency, or probability has changed or disappeared; (iii) a change is detected in the generality of the description provided by a unit of information; (iv) a previously nonexistent unit of information has been added; and (v) a unit of information is factually incorrect.

**Coding the Signal of the Message.** We also used the transcripts to analyze the signal of the conversation. Here, we reviewed all sentences uttered by the informed speaker at each chain position and labeled each of them as positive, negative, or neutral. Sentences that highlight the suspected dangers of triclosan or express a negative judgement about it received a negative label, whereas sentences suggesting that the use of triclosan is safe or well-controlled received a positive label. Sentences that were labeled neither negative nor positive were considered neutral. This procedure was conducted by two independent coders, who received the same instruction sheets (*SI Materials and Methods*). We used the mean response of the two coders. The reliability of the encoding was then assessed by comparing the number of positive and negative statements found by each coder in the same transmission. In 56% of the cases, the numbers reported by the two coders were equal. In 94% of the cases, the difference was two or less.

1. Slovic P (2000) *The Perception of Risk* (Earthscan Publications, London).
2. Slovic P (1987) Perception of risk. *Science* 236(4799):280–285.
3. Funk S, Gilad E, Watkins C, Jansen VA (2009) The spread of awareness and its impact on epidemic outbreaks. *Proc Natl Acad Sci USA* 106(16):6872–6877.
4. Huang L, et al. (2013) Effect of the Fukushima nuclear accident on the risk perception of residents near a nuclear power plant in China. *Proc Natl Acad Sci USA* 110(49):19742–19747.
5. Loewenstein GF, Weber EU, Hsee CK, Welch N (2001) Risk as feelings. *Psychol Bull* 127(2):267–286.
6. Pachur T, Hertwig R, Steinmann F (2012) How do people judge risks: Availability heuristic, affect heuristic, or both? *J Exp Psychol Appl* 18(3):314–330.
7. Slovic P, Peters E (2006) Risk perception and affect. *Curr Dir Psychol Sci* 15(6):322–325.
8. Fischhoff B, Brewer NT, Downs JT (2011) *Communicating Risks and Benefits: An Evidence-Based User's Guide* (US Food and Drug Administration, Silver Spring, MD).
9. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz L, Woloshin S (2007) Helping doctors and patients make sense of health sStatistics. *Psychol Sci Public Interest* 8(2):53–96.
10. Faria J, Krause S, Krause J (2010) Collective behavior in road crossing pedestrians: The role of social information. *Behav Ecol* 21(6):1236–1242.
11. Salathé M, Vu D, Khandelwal S, Hunter D (2013) The dynamics of health behavior sentiments on a large online social network. *European Physical Journal Data Science* 2:4.
12. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357(4):370–379.
13. Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358(21):2249–2258.
14. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197.
15. Scherer CW, Cho H (2003) A social network contagion theory of risk perception. *Risk Anal* 23(2):261–267.
16. Binder AR, Scheufele DA, Brossard D, Gunther AC (2011) Interpersonal amplification of risk? Citizen discussions and their impact on perceptions of risks and benefits of a biological research facility. *Risk Anal* 31(2):324–334.
17. Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci USA* 111(24):8788–8790.
18. Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. *Proc Natl Acad Sci USA* 108(22):9020–9025.
19. Wu F, Huberman BA (2007) Novelty and collective attention. *Proc Natl Acad Sci USA* 104(45):17599–17601.
20. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
21. Moussaïd M, Helbing D, Theraulaz G (2011) How simple rules determine pedestrian behavior and crowd disasters. *Proc Natl Acad Sci USA* 108(17):6884–6888.
22. Helbing D, et al. (2014) Saving human lives: What complexity science and information systems can contribute. *J Stat Phys* 158(3):735–781.
23. Salathé M, Khandelwal S (2011) Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLOS Comput Biol* 7(10):e1002199.
24. Moussaïd M (2013) Opinion formation and the collective dynamics of risk perception. *PLoS ONE* 8(12):e84592.
25. Renn O, Burns W, Kasperson J, Kasperson R, Slovic P (1992) The social amplification of risk: Theoretical foundations and empirical applications. *J Soc Issues* 48(4):137–160.
26. Kasperson R, et al. (1988) The social amplification of risk: A conceptual framework. *Risk Anal* 8(2):177–187.
27. Yueh M-F, et al. (2014) The commonly used antimicrobial additive triclosan is a liver tumor promoter. *Proc Natl Acad Sci USA* 111(48):17200–17205.
28. Bartlett F (1932) *Remembering: A Study in Experimental and Social Psychology* (Cambridge Univ Press, Cambridge, UK).
29. Kempe M, Mesoudi A (2014) Experimental and theoretical models of human cultural evolution. *WIREs Cognitive Science* 5(3):317–326.
30. Griffiths TL, Kalish ML, Lewandowsky S (2008) Review. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philos Trans R Soc Lond B Biol Sci* 363(1509):3503–3514.
31. Horner V, Whiten A, Flynn E, de Waal FB (2006) Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc Natl Acad Sci USA* 103(37):13878–13883.
32. Kirby S, Cornish H, Smith K (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc Natl Acad Sci USA* 105(31):10681–10686.
33. Mesoudi A, Whiten A, Laland KN (2006) Towards a unified science of cultural evolution. *Behav Brain Sci* 29(4):329–347, discussion 347–383.
34. Soll JB, Larrick RP (2009) Strategies for revising judgment: How (and how well) people use others' opinions. *J Exp Psychol Learn Mem Cogn* 35(3):780–805.
35. Moussaïd M, Kämmer JE, Analytis PP, Neth H (2013) Social influence and the collective dynamics of opinion formation. *PLoS ONE* 8(11):e78433.
36. Festinger L (1957) *A Theory of Cognitive Dissonance* (Stanford Univ Press, Stanford, CA).
37. Lorenz J (2007) Continuous opinion dynamics under bounded confidence: A survey. *Int J Mod Phys* 18(12):1819–1838.
38. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442.
39. Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proc Natl Acad Sci USA* 105(41):15649–15653.
40. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81(2):591–646.
41. Arkes HR, Gaissmaier W (2012) Psychological research and the prostate-cancer screening controversy. *Psychol Sci* 23(6):547–553.